Persistent topology of the reionization bubble network - I. Formalism and phenomenology

Elbers, Willem; van de Weygaert, Rien

Published in:
Monthly Notices of the Royal Astronomical Society

DOI:
10.1093/mnras/stz908

Publication date:
2019

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **486,** 1523–1538 (2019)
Advance Access publication 2019 March 30

doi:10.1093/mnras/stz908

Persistent topology of the reionization bubble network – I. Formalism and phenomenology

Willem Elbers ★ and Rien van de Weygaert
Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700AV Groningen, the Netherlands

Accepted 2019 March 25. Received 2019 March 1; in original form 2018 December 2

ABSTRACT

We present a new formalism for studying the topology of H II regions during the Epoch of Reionization, based on persistent homology theory. With persistent homology, it is possible to follow the evolution of topological features over time. We introduce the notion of a persistence field as a statistical summary of persistence data and we show how these fields can be used to identify different stages of reionization. We identify two new stages common to all bubble ionization scenarios. Following an initial pre-overlap and subsequent overlap stage, the topology is first dominated by neutral filaments (filament stage) and then by enclosed patches of neutral hydrogen undergoing outside-in ionization (patch stage). We study how these stages are affected by the degree of galaxy clustering. We also show how persistence fields can be used to study other properties of the ionization topology, such as the bubble size distribution and the fractal-like topology of the largest ionized region.

Key words: intergalactic medium – cosmology: theory – dark ages, reionization, first stars.

1 INTRODUCTION

The Epoch of Reionization was a cosmic phase transition in which the neutral hydrogen of the post-recombination era was ionized by the first luminous objects. Reionization coincides with and influences the formation of the first galaxies, resulting in a complex and non-linearly evolving ionization fraction field $x_{\rm II} = N_{\rm II}/(N_{\rm I} + N_{\rm II})$. The topology of this ionization field has been the subject of sustained theoretical interest. One hope is that the topology will tell us about the physical processes involved and in particular about the sources responsible for reionization (Friedrich et al. 2011; Katz et al. 2018). With currently ongoing observations of the redshifted 21-cm line (Beardsley et al. 2016; Patil et al. 2017; Kerrigan et al. 2018), we will for the first time gain access to statistics of the 21-cm field and the closely related ionization field. If techniques improve sufficiently, we will even be able to image the ionization field through 21-cm tomography, which is one of the goals of the Square Kilometre Array (Mellema et al. 2015). To connect these observations to the many simulations[1] of the reionization era, it is important to develop robust measures that capture a sufficient level of detail of the ionization topology and are appropriate for every stage of the ionization process. This study is an effort to develop such a measure by borrowing from the theory of persistent homology. In this first paper of two, we explain our methodology and illustrate the usefulness of persistent homology with a number of phenomenological models. In a follow-up paper, we apply these ideas to more realistic scenarios.

1.1 Topology of reionization

An early qualitative description of the topology of reionization goes back to Gnedin (2000), who identified three stages of reionization. During the *pre-overlap stage*, radiation emitted by the first luminous objects ionizes the dense surrounding gas, forming localized bubbles of ionized material. These bubbles then expand into the low-density intergalactic medium. In a second *overlap stage*, the ionized regions merge and the global ionization fraction rises rapidly. Finally, in the *post-overlap stage*, the remaining high-density neutral pockets are ionized from the outside. This picture of reionization can be described in terms of *inside-out* and *outside-in* reionization (Lee et al. 2008; Choudhury et al. 2009; Friedrich et al. 2011). These terms refer to the ionization of high-density regions: high-density regions containing ionizing sources are ionized first and their bubbles expand outward (inside-out), but high-density regions without ionizing sources are ionized from the outside at the end of reionization (outside-in). Rather than high-density pockets, high-density filaments might also be the last regions to be ionized (Finlator et al. 2009). Either way, the degree to which outside-in reionization occurs depends on the minimum halo mass necessary for ionizing sources to form, demonstrating one way in which the topology reflects the underlying physics. Another example is the

★ E-mail: elbers@astro.rug.nl

[1] State of the art simulations include Gnedin (2014), Iliev et al. (2014), Ocvirk et al. (2016), Pawlik et al. (2017), and Doussot, Trac & Cen (2019). Seminumerical approximations are also commonly used (Mesinger & Furlanetto 2007; Choudhury, Haehnelt & Regan 2009; Mesinger, Furlanetto & Cen 2011; Zahn et al. 2011; Majumdar et al. 2014; Hutter 2018).

degree of galaxy clustering, which affects the patchiness of the ionization field (Iliev et al. 2014).

The reionization process has been most commonly quantified with the 21-cm power spectrum or more directly with the power spectrum of the ionization field, which constitutes the dominant component of the 21-cm power spectrum during the latter half of reionization (Iliev et al. 2014). The 21-cm power spectrum is the first observable that is likely to be measured and contains valuable information. The overall amplitude of the 21-cm power spectrum tracks the progress of reionization, since the differential brightness temperature is proportional to the fraction of neutral hydrogen.[2] The amplitude of the ionization power spectrum peaks in the middle of reionization when variance in the ionization field is highest (Hutter 2018). A general finding is that once reionization has started, the ionization power spectrum peaks at some scale indicative of a characteristic bubble size, which increases as reionization progresses and bubbles merge (Furlanetto, Zaldarriaga & Hernquist 2004b; Zahn et al. 2011; Hong et al. 2014; Iliev et al. 2014; Majumdar et al. 2014; Dixon et al. 2016; Hutter 2018). The power spectrum can also be used to identify more complex patterns in the ionization topology. Friedrich et al. (2011) found two peaks and explained this with two periods of ionization bubble formation interceded by a period of suppression. The slope of the power spectrum may indicate to what degree ionization occurred outside-in (Choudhury et al. 2009). Finally, the 21-cm power spectrum also carries information on pre-reionization physics (Mesinger et al. 2011). Nevertheless, the power spectrum is not enough to characterize the evidently non-Gaussian ionization field. Kakiichi et al. (2017) nicely demonstrated that the 21-cm signal from a radiative transfer simulation is morphologically very different from a Gaussian random field with the same power spectrum. Hence, complementary observables such as the bispectrum (Shimabukuro et al. 2017) are needed (this paper introduces another such observable).

A common alternative has been to study the morphology of individual ionization bubbles. Many authors have looked at the size distribution of ionization bubbles (Furlanetto, Hernquist & Zaldarriaga 2004a; McQuinn et al. 2007; Mesinger & Furlanetto 2007; Friedrich et al. 2011; Zahn et al. 2011; Malloy & Lidz 2013; Lin et al. 2016; Giri et al. 2017; Kakiichi et al. 2017) or at the shape of such bubbles (Gleser et al. 2006; Iliev et al. 2006; Furlanetto & Oh 2016; Bag et al. 2018; Kapahtia et al. 2018). They typically find that the bubble radius is approximately lognormally distributed with a characteristic scale that increases and a variance that decreases as reionization progresses.

Recently, reionization has also been fruitfully studied from the perspective of percolation theory (Furlanetto & Oh 2016; Bag et al. 2018). A salient feature of the qualitative description above is the sharp rise in ionization fraction during the overlap stage. This can be understood as a phase transition associated with the percolation of ionization bubbles. The transition is characterized by the appearance of one large connected cluster of ionized regions that spans the simulation box. Near the phase transition, the ionized regions demonstrate the scaling behaviour expected from universality.

[2]Indeed, we have (Pritchard & Loeb 2012):

$$\delta T_{21}(z) = T_0(z)(1+\delta_b)(1-x_{\rm II})\left(1-\frac{T_{\rm CMB}(z)}{T_{\rm S}}\right),$$

where $T_{\rm S}$ is the spin temperature, $T_0(z)$ a function of cosmological parameters and redshift z, and δ_b the baryonic overdensiy.

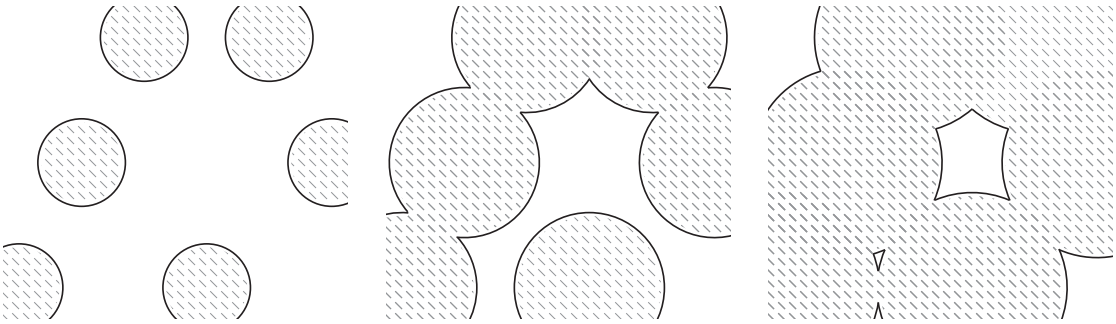
In terms of purely topological measures, the most basic is probably the number k of connected components. Starting from a discretized snapshot of the ionization field, this can be determined by applying a friends-of-friends algorithm (Friedrich et al. 2011), watershed algorithm (Platen, van de Weygaert & Jones 2007; Lin et al. 2016) or technique called granulometry (Kakiichi et al. 2017) to the points with an ionization fraction above a certain threshold. The evolution of the number of ionized regions alone can already tell the qualitative story of emerging and then rapidly merging bubbles (Fig. 1). A more detailed variation is to follow the evolution of individual ionized regions and to construct merger trees, which allows one to study the number density of new, expanding, and merging regions over time (Chardin, Aubert & Ocvirk 2012).

Another elementary topological property is the genus g, which is the number of cuts one can make without increasing the number of components, or the related Euler characteristic $\chi = 2k - 2g$. More complex still, the Minkowski functionals combine geometric properties such as the volume, surface area, and mean curvature of the ionized region with the Euler characteristic. Both genus and Minkowski functionals have been applied in this context. Different stages of reionization can be distinguished by means of genus curves (Lee et al. 2008) and Minkowski functionals (Gleser et al. 2006). Both can be used to constrain various source properties (Friedrich et al. 2011). The 21-cm field too has been studied with genus curves (Hong et al. 2014) and Minkowski functionals (Yoshiura et al. 2017), both agreeing that they can be used to constrain physics if accurate images of the 21-cm signal were available. Kapahtia et al. (2018) used Minkowski tensors to characterize the size and shape distribution of ionization bubbles and Bag et al. (2018) used ratios of Minkowski functionals called *shapefinders* (Sheth et al. 2003; Shandarin, Sheth & Sahni 2004) to express such properties as the length, thickness, and breadth of the largest ionized region. They found that the largest ionized region that emerges during the phase transition has a complex and highly filamentary topology.

To summarize, most studies thus far have focused on global features of the topology such as the Euler characteristic or on the morphology of individual ionization bubbles. However, the picture of disconnected ionization bubbles is only appropriate during the pre-overlap stage when the global ionization fraction is relatively small. During most of reionization, most of the ionized material is contained in one connected structure that stretches the length of the Universe and has a complicated and fractal-like topology (Furlanetto & Oh 2016; Bag et al. 2018). We would therefore like to find tools that help us understand the topology during the later stages of reionization, especially since the later stages are easiest to observe through the 21-cm signal.

1.2 Persistent homology

In this paper, we show that persistent homology is ideally suited to study the process of cosmic reionization through its topology. As a subfield of mathematics, topology is concerned with properties that are preserved under continuous deformations (like bending or stretching). An important example of such a property is the number of holes. Counting holes is therefore a useful way to distinguish topologies. In Fig. 2, we see three examples of holes in different dimensions. A zero-dimensional hole is a gap that separates a connected component, like a distinct H II region, from the space surrounding it. There is one gap for each component, so we often blur the distinction. A one-dimensional hole is an opening like the cross-section of a tunnel. Finally, a two-dimensional hole is a



(a) $\alpha = 0.7$ (b) $\alpha = 1.2$ (c) $\alpha = 1.5$

Figure 1. Filtration of a uniformly sized bubble network.

cavity or void surrounded by a shell. We will generally refer to gaps, tunnels, and voids as *topological features*.

We are interested in topological features in the ionization field. The connected components of this field are simply the ionization bubbles or the distinct H II regions. The tunnels are neutral filaments that pierce through the ionization bubble network. Voids are patches of neutral hydrogen enclosed by ionized material. We stress that these are voids in the ionization field, often corresponding to overdensities and distinct from cosmological voids, which correspond to underdensities. We refer to these features as ionized bubbles ($k = 0$), neutral filaments or tunnels ($k = 1$), and neutral patches ($k = 2$).

In homology theory, we count holes by classifying the loops that can be drawn on an object. This is possible because of a correspondence between loops and holes. We are primarily interested in the so-called Betti numbers. Formally, the kth Betti number β_k is the rank of the kth homology group, which contains the distinct classes of k-dimensional loops. Intuitively, the kth Betti number is simply the number of k-dimensional holes. In other words, the zeroth Betti number β_0 describes the number of connected components, the first Betti number β_1 the number of tunnels, and the second Betti number β_2 the number of voids.

Together, the Betti numbers contain strictly more information than the Euler characteristic $\chi = \beta_0 - \beta_1 + \beta_2$. As the number of ionized regions is initially much larger than the number of enclosed filaments and neutral patches, the Euler characteristic has sometimes been thought of as a measure of the number of bubbles: $\chi \approx \beta_0$. However, it is interesting to consider the Betti numbers separately. We should for instance be able to see the filamentary nature of reionization by looking at β_1. Neutral patches undergoing outside-in ionization can be identified by looking at β_2.

We can go one step further by taking topological persistence into account (Edelsbrunner, Letscher & Zomorodian 2000; Zomorodian & Carlsson 2005). Rather than dealing with a static object, we consider a nested sequence of objects[3] called a filtration. It facilitates a formal mathematical description of the hierarchical evolution of structure. Intuitively, we picture a filtration as an expanding bubble network, as depicted in Fig. 1. Each element in the sequence is assigned a scale α. By studying the topology at every scale, we compute a birth date $\alpha_{\rm birth}$ and death date $\alpha_{\rm death}$ for all topological features. In Fig. 1, we see the death of multiple gaps and the birth of two tunnels. The difference $\alpha_{\rm death} - \alpha_{\rm birth}$ is the *persistence* of a feature. In a persistence diagram, all features are plotted in the ($\alpha_{\rm birth}$, $\alpha_{\rm death}$) plane. Persistence diagrams contain even more information

[3]In which each element of the sequence contains the previous element.

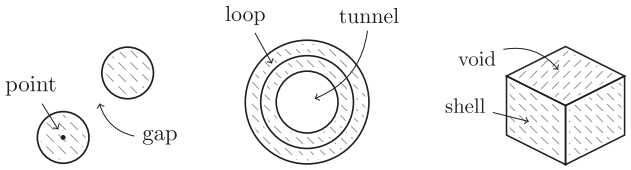


Figure 2. The homology of an object refers to the distinct classes of loops that can be drawn on it, or equivalently about its boundaries and holes. A k-dimensional loop (point, loop, shell) can be continuously deformed until it meets a k-dimensional hole (gap, tunnel, void). Shown are $k = 0, 1, 2$.

than Betti numbers, which only count the numbers of topological features at a given scale. For example, if we consider the filtration of a bubble network along the time axis, we can see not just the number of neutral patches but also how long it takes for them to be ionized.

In the context of reionization, there are three interesting dimensions along which to build a filtration.

(i) *Time.* The most straightforward interpretation is to imagine the filtration as a bubble network evolving over time. In this case, $\alpha_{\rm birth}$ and $\alpha_{\rm death}$ are literally the birth and death dates of topological features. The persistence is simply the lifetime of a feature. A temporal filtration shows the hierarchical build-up of structure.

(ii) *Space.* Given a time slice of the ionization history, we can also probe the connectivity structure of the bubble network. In this case, $\alpha_{\rm birth}$ and $\alpha_{\rm death}$ refer to spatial scales at which features arise. The persistence is now a measure of the topological significance of a feature. A spatial filtration looks into the multiscale structure that emerges as a result of hierarchical evolution.

(iii) *Ionization fraction.* In this paper, we assume a binary ionization field. However, we can also construct a filtration by lowering the ionization threshold (the ionization fraction above which a point is considered ionized). The persistence of a feature is now interpreted as the differential ionization fraction of the hole. For instance, the persistence of an opening tells us about the ionization state of the enclosed filament. In this study, we consider only filtrations along the first two dimensions.

With developments in computational topology over the past two decades, persistent homology is now readily applicable in various practical situations. It has become the preeminent tool of topological data analysis (Zomorodian 2012; Wasserman 2018). In cosmology, persistent homology has previously been applied to the cosmic web (Sousbie 2011; van de Weygaert et al. 2011; Nevenzeel 2013; Pranav et al. 2017; Xu et al. 2019), to Gaussian random fields (Feldbrugge & van Engelen 2012; Park et al. 2013; Cole & Shiu 2018; Feldbrugge

et al. 2018; Pranav et al. 2018), and to interstellar magnetic fields (Makarenko et al. 2018).

We further discuss the theory of filtrations and homology in Section 2. In Section 3, we describe our methodology and elaborate on the interpretation of bubble network filtrations. In Section 4, we discuss how the bubble network depends on the properties and spatial distribution of ionizing sources. The interpretation of persistent homology is explained in Section 5 using a number of phenomenological models. Finally, we conclude in Section 6.

2 THEORY

Our formalism makes use of persistent homology theory to analyse bubble networks. We also borrow a tool from computational topology called α-shapes to model these bubble networks. The first part of this section deals with α-shapes and its weighted generalization. The latter is needed to model non-uniform bubble networks. We discuss an alternative to α-shape filtrations in Section 2.3. The rest of the section is concerned with homology theory, topological persistence, and the statistics of persistence diagrams.

2.1 α-shapes

Homology groups and the associated Betti numbers are most easily computed for a class of relatively simple objects called simplicial complexes. A simplicial complex is a structure built from points, lines, triangles, and higher dimensional analogues called *simplices*.[4] An illustration of a simplicial complex is shown in the second panel of Fig. 3. Of particular interest is the idea of a *filtration* of a simplicial complex $\mathcal{K}$. This is a nested sequence of simplicial complexes $\varnothing \subseteq \mathcal{K}^0 \subseteq \mathcal{K}^1 \subseteq \cdots \subseteq \mathcal{K}^m = \mathcal{K}$. By computing the homology at each step, we can follow how the topology changes as points, lines, and triangles are filled in. There are different ways to translate the complex reionization topology into a usable filtration. The most straightforward way to accomplish this task is with α-shapes (Edelsbrunner, Kirkpatrick & Seidel 1983; Edelsbrunner & Mücke 1994).

α-shapes are families of geometric constructions that capture the shape of a point set $\mathcal{P}$ over a range of scales. In this paper, we take as our point set the collection of bubble centres. The α-shape is then constructed as follows. We start with the *Voronoi tesselation* of $\mathcal{P}$ (Icke & van de Weygaert 1987; Okabe 1992; van de Weygaert 1994). This is a partition of $\mathbb{R}^3$ into cells, one for each point $p \in \mathcal{P}$. The Voronoi cell of p consists of those points $x \in \mathbb{R}^3$ that are at least as close to p as to any other point $q \in \mathcal{P}$. The *Delaunay triangulation* $\mathcal{T}$ of $\mathcal{P}$ is the dual graph of the Voronoi tesselation. Two points in $\mathcal{P}$ are connected by an edge in $\mathcal{T}$ if their Voronoi cells intersect. The Delaunay triangulation $\mathcal{T}$ is a simplicial complex. Its simplices are spanned by the sets of $k+1$ points in $\mathcal{P}$ whose circumscribing sphere does not contain any other point in $\mathcal{P}$. See the first two panels in Fig. 3 for an example. By taking suitable subsets of $\mathcal{T}$, we get a filtration.

For any value of $\alpha \geq 0$, we define the *α-complex* as a particular subset of the Delaunay triangulation. We draw a ball of radius[5] α around each point $p \in \mathcal{P}$. Those simplices of $\mathcal{T}$ that are contained within the union of balls belong to the α-complex. The α-shape is the union of all simplices in the α-complex. As α grows larger, the α-shape gets filled in. This is what we see in the last two panels of Fig. 3. The α-shape only changes at discrete values of α. By increasing α until the entire Delaunay triangulation is filled in, we produce our desired filtration.

A crucial point is that the bubble network, which we now understand as the union of all closed balls of radius α centred on a point in $\mathcal{P}$, is *homotopy equivalent* to the corresponding α-shape. This condition is slightly weaker than being *homeomorphic*, in which case all topological properties of the two shapes would be identical, but it does mean that the shapes can be continuously deformed into each other. In particular, it implies that the bubble network and the α-shape have the same number of holes, validating our approach. Of course, the same results are valid for any other shape in the homotopy class, which includes more realistically shaped bubble networks obtained by deforming the spherical network.

2.2 Weighted α-shapes

To model ionization bubbles of different sizes or born at different times, we need to go beyond the simple α-shapes of the previous section. In this case, *weighted α-shapes* provide the appropriate filtration set (Edelsbrunner 1992; Edelsbrunner et al. 1995). This is a generalization of the above construction, where each point p is assigned a weight w_p. We picture a weighted point (p, w_p) as a sphere centred on p with radius w_p. Consider the weighted point set $\mathcal{P}$. Let $\mathcal{B}$ be the set of closed balls with boundary in $\mathcal{P}$. The union $\mathcal{F} = \bigcup \mathcal{B}$ of these balls is what we understand as a bubble network.

Define the *weighted distance* from (p, w_p) to (q, w_q) as

$$\pi(p, q) = \|p - q\|^2 - w_p^2 - w_q^2, \tag{1}$$

where $\|p - q\|$ is the Euclidean distance. The *weighted Voronoi cell* of (p, w_p) consists of all unweighted points $x \in \mathbb{R}^3$ whose weighted distance to p is no more than the weighted distance to any other $q \in \mathcal{P}$. The *weighted Delaunay triangulation* is then the dual graph of the *weighted Voronoi tesselation*.

Denote by $\mathcal{F}_\alpha$ the bubble network that is obtained by inflating every sphere (p, w_p) to a sphere (p, r) with radius

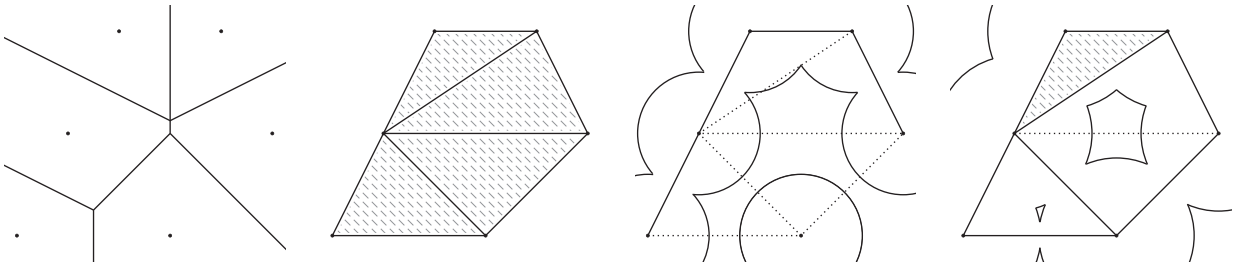
$$r = \sqrt{\mathrm{sign}(w_p) w_p^2 + \mathrm{sign}(\alpha)\alpha^2}. \tag{2}$$

We explicitly allow for negative values of α, so that we can both inflate ($\alpha > 0$) and deflate ($\alpha < 0$) the bubbles. The interpretation of negative weights ($w_p < 0$) is explained later. Points with $r^2 < 0$ are called *redundant* and are omitted. The reason for using the non-linear radius function (2) is that the resulting Voronoi cells are unchanged when α is varied. It follows that the dual Delaunay triangulations are also independent of α, allowing us to build a filtration analogous to the unweighted case.

The weighted α-complex is constructed as follows. Let $\sigma \in \mathcal{T}$ be a simplex in the weighted Delaunay triangulation. The simplex is part of the α-complex if it is the face of another simplex in the complex or if it is 'smaller than α'. This agrees with our intuition for the unweighted case, where a simplex was added as soon as the balls were large enough to contain it, but the formal definition requires some thought. We define the *size* y_σ of σ to be the smallest value of α for which the α-inflated spheres centred on its $k+1$ vertices intersect in a point x. Equivalently, y_σ is the radius of the smallest sphere x, such that $\pi(x, p) = 0$ for all vertices p of σ. It may be useful to note that $\pi(x, p) = 0$ if and only if the spheres x

[4] Technically, a *k-simplex* σ is the smallest convex set that contains its $k+1$ affinely independent vertices. A *face* of σ is any simplex spanned by a subset of its vertices. A *simplicial complex* $\mathcal{K}$ is any set of simplices such that if $\sigma \in \mathcal{K}$ is a simplex, then the faces of σ also belong to $\mathcal{K}$ and such that any two simplices in $\mathcal{K}$ are either disjoint or intersect in a common face.

[5] Another commonly used convention is that the radius of the ball is $\sqrt{\alpha}$.



(a) Voronoi tesselation (b) Delaunay triangulation (c) Alpha-shape for $\alpha = 1.2$ (d) Alpha-shape for $\alpha = 1.5$

Figure 3. The idea behind simplicial homology is that we can study the homology of a complex object by looking at the homology of an associated structure of points, lines, and triangles (simplices), which are computationally easier to handle. In the final panel, notice that the bottom triangle is not filled in, correctly capturing the opening that exists in the bubble network. Compare Fig. 1.

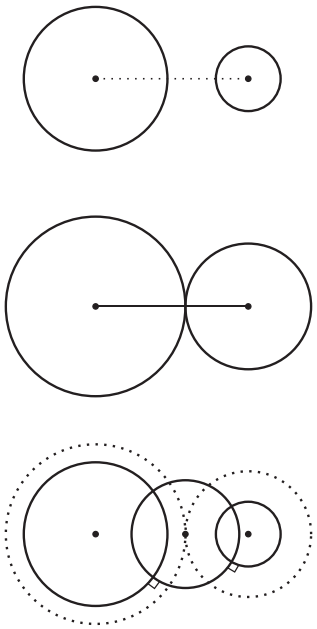


Figure 4. Two weighted points and the simplex σ spanned by their centres (top). The spheres are α-inflated via equation (2) until they intersect, at which point the edge enters the α-complex (middle). If we place a sphere of radius α at the point of intersection, then it is orthogonal to the original uninflated spheres (bottom). The size y_σ of the edge is α.

and p are orthogonal. Now we say that the simplex is part of the α-complex if $\alpha \geq y_\sigma$, provided there are no conflicts with other points in $\mathcal{P}$. A conflict occurs if $\pi(x, q) < 0$ for any point $q \in \mathcal{P}$ that is not a vertex of σ. See Fig. 4 for an example where σ is a line segment.

We return to the point on negative weights. If we have bubbles at locations $\{p_1, p_2, \dots\}$ born at times $\{\tau_1, \tau_2, \dots\}$, we define the weights by $w_i = -\tau_i$. This ensures that for $\alpha < \tau_i$, the point p_i is redundant, but for $\alpha \geq \tau_i$, we get an inflating bubble with radius $\sqrt{\alpha^2 - \tau_i^2}$. A point with negative weight $w_p < 0$ is thus interpreted as a bubble born at 'time' $\alpha = -w_p > 0$.

2.3 Field filtrations

Instead of using α-shapes, we may wish to build a filtration that directly reflects the properties of some scalar field $f : \mathbb{R}^3 \to \mathbb{R}$, such as the ionization fraction field. One way to do this is as follows. We first specify a point set $\mathcal{P}$ at which the field is sampled, such that each vertex $p \in \mathcal{P}$ has an associated field value $f(p)$. As in the α-shape method, the filtration consists of subsets of a Delaunay triangulation of $\mathcal{P}$. At any value of $\alpha \in \mathbb{R}$, those vertices p with $f(p) \geq \alpha$, as well as any simplices connecting them, are part of the simplicial complex $\mathcal{K}_\alpha$. Assuming that f is smooth, $\mathcal{K}_\alpha$ only changes when α passes through a critical value of the field f. We thus obtain our desired filtration $\varnothing \subseteq \mathcal{K}^0 \subseteq \mathcal{K}^1 \subseteq \cdots \subseteq \mathcal{K}^m = \mathcal{K}$.

An important question concerns how to choose the point set $\mathcal{P}$ in a way that preserves the topology of the field. When we have discrete samples or measurements of the field, this can be done with the Delaunay Tessellation Field Estimator DTFE (Schaap & van de Weygaert 2000; van de Weygaert & Schaap 2008; Cautun & van de Weygaert 2011). This method has previously been applied to the cosmic density field by Pranav et al. (2017). We refer to their paper for more details on this approach.

2.4 Homology

Betti numbers are derived from the field of algebraic topology. Algebraic topology is about finding ways of mapping topological spaces to algebraic objects, such as groups. One example, and the one in which we are interested, is that of homology groups. As mentioned before, the idea behind homology is that we can characterize the topology of an object in terms of the cycles or loops that we can draw on it. Equivalently, homology tells us about the boundaries of and holes in a space. Two loops are equivalent when they can be continuously deformed into each other. On the sphere any loop can be contracted to a point, but on the torus there are two classes of non-contractible loops that cannot be deformed into each other. This corresponds to the fact that there are no one-dimensional holes in the sphere and two distinct one-dimensional holes in the torus: the hole through the middle and any cross-section of the tunnel that runs along the torus.

We can generalize the idea of loops and holes to arbitrary dimensions (see Fig. 2). As previously explained, we have points and gaps ($k = 0$), loops and tunnels ($k = 1$), and shells and voids ($k = 2$). In arbitrary dimensions, we talk about k-cycles surrounding k-dimensional holes. The homology classes in dimension k can be arranged into a group called the kth homology group $\mathcal{H}_k$. The kth Betti number β_k is the rank of this group. We arrive again at the notion that the kth Betti number describes the number of k-dimensional holes. In Section 2.5, we give a little more insight in how these notions are defined for simplicial complexes. Refer to Munkres (1984) and Hatcher (2002) for a textbook introduction.

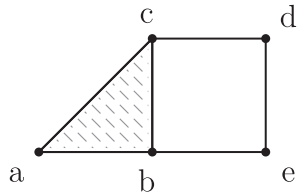


Figure 5. A simplicial complex consisting of five points, six line segments, and one (filled in) triangle.

2.5 Simplicial homology

The homology of a simplicial complex can be defined in terms of chains of simplices. To demonstrate this, consider the simplicial complex in Fig. 5. There are five 0-simplices, namely the points a, $\ldots$, e. There are six 1-simplices or line segments, which we write as $[a, b]$. There is one 2-simplex, namely the triangle $[a, b, c]$. We start with the observation that these simplices can be chained together. For instance, we could write the path around the triangle as $\sigma = [a, b] + [b, c] + [c, a]$. In general, we call any linear combination of k-simplices with integer coefficients (modulo p) a k-chain. With the operation of addition, the k-chains form a free Abelian group[6] called the kth chain group $\mathcal{C}_k$.

Given a k-chain σ, we can construct a $(k-1)$-chain $\partial\sigma$ called its *boundary*. For instance, the boundary of a line segment is the difference of its endpoints and the boundary of a triangle is the path around it. A chain whose boundary is zero is called a *cycle*. The boundary of the 2-chain $[a, b, c]$ is the 1-chain $\sigma = [a, b] + [b, c] + [c, a]$. The boundary of σ is 0, since its endpoints coincide. Hence, σ is also a 1-cycle. The path τ around the square is similarly a 1-chain and a 1-cycle, but not a boundary since it does not enclose any triangles. The boundaries and cycles form subgroups of the chain group, denoted as $\mathcal{B}_k$ and $\mathcal{Z}_k$ respectively.

Two cycles are *homologous* if they differ by a boundary. Visually, this means they surround the same holes. For example, the cycle $\sigma + \tau$ that encircles the combined triangle and square figure is homologous to the cycle τ that just goes round the square, because the difference σ is a boundary. Being homologous is an equivalence relation. All k-cycles can thus be partitioned into homology classes. These homology classes form a group called the kth homology group $\mathcal{H}_k$. This can also be understood as the factor group $\mathcal{H}_k = \mathcal{Z}_k/\mathcal{B}_k$. Recall that the kth Betti number is the rank of $\mathcal{H}_k$. In the example above, there are two independent 1-cycles namely the path σ around the triangle and the path τ around the square. Thus the 1-cycle group $\mathcal{Z}_1$ has a basis $\{\sigma, \tau\}$ and rank 2. There is only one independent 1-boundary, namely σ, so the 1-boundary group $\mathcal{B}_1$ has rank 1. Hence, we find that $\beta_1 = \text{rank}\,\mathcal{H}_1 = \text{rank}\,\mathcal{Z}_1 - \text{rank}\,\mathcal{B}_1 = 1$. Intuitively, this agrees with the fact that there is one one-dimensional hole, namely the one enclosed by the square.

2.6 Persistence diagrams

As the previous example shows, the problem of identifying the k-dimensional holes in a simplicial complex can be solved by finding suitable bases for the cycle and boundary groups $\mathcal{Z}_k$ and $\mathcal{B}_k$. Computationally, it is convenient to do this by representing the boundary operator as a matrix. As an example, consider a simplicial complex consisting of a single triangle $[a, b, c]$ with boundary $\sigma = [a, b] + [b, c] + [c, a]$. Recall that the boundary of an edge is the difference of its endpoints: $\partial[a, b] = b - a$. With respect to the basis $\{a, b, c\}$, we write the boundaries of the 1-simplices as

$$\begin{bmatrix} & [a,b] & [b,c] & [c,a] \\ a & -1 & 0 & 1 \\ b & 1 & -1 & 0 \\ c & 0 & 1 & -1 \end{bmatrix} \sim \begin{bmatrix} & [a,b] & [b,c] & \sigma \\ b-a & 1 & 0 & 0 \\ c-b & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

On the right, we have brought the matrix to Smith normal form by means of elementary row and column operations (switching rows or columns, multiplying them by a non-zero scalar, and adding multiples of one row or column to another). In this form, we see that $\{b - a, c - b\}$ is a basis for the boundary group $\mathcal{B}_0$ and $\{\sigma\}$ is a basis for the cycle group $\mathcal{Z}_1$.

We use similar techniques to compute the persistent homology of a filtration $\varnothing \subseteq \mathcal{K}^0 \subseteq \mathcal{K}^1 \subseteq \cdots \subseteq \mathcal{K}^m = \mathcal{K}$. The goal is to identify every hole that appears in the filtration. Each hole first appears as a cycle in some complex $\mathcal{K}^i$. If the hole is still present in $\mathcal{K}$, we assign it a pair (i, ∞). Other holes disappear when they are filled up. This occurs when the corresponding cycle becomes a boundary, say in $\mathcal{K}^j \supseteq \mathcal{K}^i$. In that case, we assign the hole a pair (i, j).

To compute these (birth, death)-pairs, we use the algorithm of Zomorodian & Carlsson (2005). The basic idea is as follows. We loop through every simplex σ^j in the order in which they appear in the filtration. We maintain matrices akin to the ones above, but suitably generalized to represent the homology of an entire filtration. The matrices are updated incrementally using elementary column operations. For each simplex σ^j, we first compute its boundary. We then check if the boundary corresponds to a zero column in the boundary matrix. If not, we find the simplex σ^i in the boundary that appears latest in the filtration. The set-up then guarantees the existence of a cycle that is born when σ^i enters the filtration and becomes a boundary when σ^j is added. The corresponding hole is assigned the pair (i, j). A more extensive discussion of a very similar computational paradigm can be found in Pranav et al. (2017).

We have implemented this algorithm for α-shape filtrations.[7] Let us make a few practical remarks. The algorithm computes the persistent homology over a finite field F_p. This is why we mentioned above that the coefficients of the chains are integers modulo a prime number p. We used $p = 2$ for the results in this paper. The algorithm also works for larger p, but the differences are negligible for our purposes. Secondly, the algorithm outputs a pair (i, j) for each feature, corresponding to the indices of the complexes in which the feature first appears and disappears, respectively. Since each complex $\mathcal{K}^i$ has an associated scale $\alpha^i \in \mathbb{R}$, this is equivalent to computing the $(\alpha_{\text{birth}}, \alpha_{\text{death}})$ values. Finally, we note that the algorithm works with an efficient data structure. This means we do not actually maintain the boundary matrices, which would be impractical for our application.

Having successfully computed the (birth, death)-pairs, we can plot the topological features in $(\alpha_{\text{birth}}, \alpha_{\text{death}})$-space, producing a persistence diagram. See Fig. 6 for an example. The horizontal (or vertical) distance of a point to the diagonal is its persistence. This particular example shows the births and deaths of tunnels in a clustered model. The persistence diagram reflects the topology of the model. We see for instance that there are two generations of features: a large number of low-persistence features on small scales and a small number of high-persistence features on large scales.

[6] Every k-chain in $\mathcal{C}_k$ is a formal sum of elements of a basis B, consisting of the k-simplices in the complex. We say that the group is free over B.

[7] Our software is available at http://willemelbers.com/persistent-homology/.

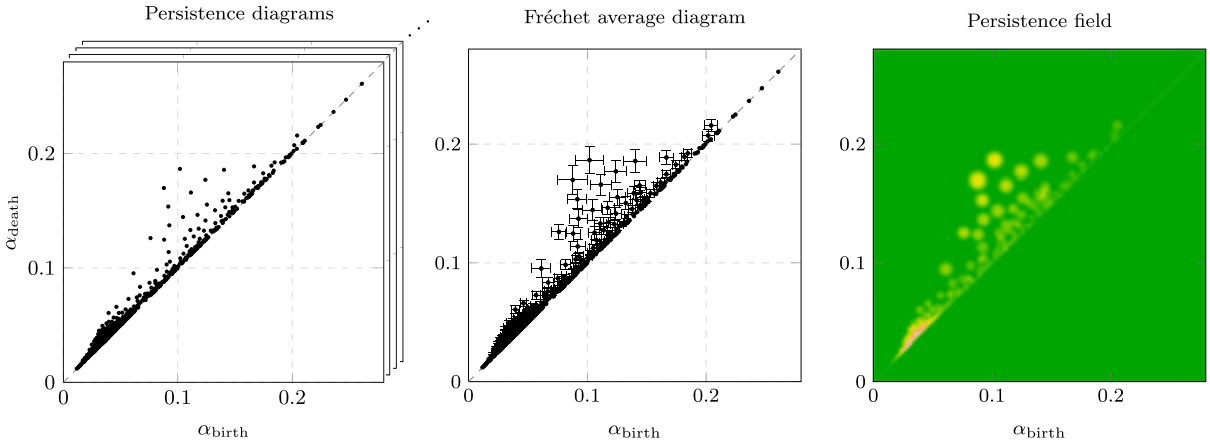


Figure 6. Our pipeline for creating persistence fields. Starting with n realizations of a stochastic process, we obtain a sample of n persistence diagrams $\{X_i\}$. We compute a Fréchet average diagram Y and associated variances σ_y^2 of the points $y \in Y$. These are then used to produce a persistence field.

The former correspond to mergers within clusters and the latter to mergers of clusters. See Section 5.1.2 for a detailed discussion of this model.

2.7 Statistics of persistence diagrams

If the preceding theory is to be applied to real world data, we must be able to handle experimental uncertainties. Even when dealing with simulations, a statistical approach is highly preferable. In this paper, the set-up is as follows. For each of the phenomenological models treated in Section 5, we generate n random bubble networks and compute one persistence diagram X_i for each realization $i = 1, \ldots, n$. We are looking for appropriate summary statistics of the sample $S = \{X_i\}$.

To describe the homology from a statistical point of view, we therefore consider the space $\mathcal{D}$ of persistence diagrams. A persistence diagram is nothing more than a collection of (α_{birth}, α_{death})-pairs, but we need an additional technical condition to ensure that $\mathcal{D}$ is a well-behaved probability space. Formally then, we define a *persistence diagram* as a countable set of finitely many points $x \in \mathbb{R}^2$ together with infinitely many copies of the diagonal $\Delta = \{(x, y) \in \mathbb{R}^2 \mid x = y\}$. In that case, $\mathcal{D}$ is a complete and separable metric space on which probability measures, expectation values, and variances can be defined (Mileyko, Mukherjee & Harer 2011). In this paper, we use the L^2-Wasserstein metric (Turner et al. 2014):

$$d(X, Y) = \left[\inf_{\phi: X \to Y} \sum_{x \in X} \|x - \phi(x)\|^2\right]^{1/2}. \tag{3}$$

To compute the distance between two persistence diagrams $X, Y \in \mathcal{D}$, we need to consider all bijections ϕ: $X \to Y$. These are one-to-one maps that match each point $x \in X$ with a point $y \in Y$ and vice versa. Here, we treat the diagonal Δ as a point that can be matched either with an off-diagonal point or with another copy of the diagonal. Given such a matching ϕ, the distance $\|x - \phi(x)\|$ is simply the Euclidean distance from x to its partner $\phi(x)$. We specify that the distance x-Δ is the distance from x to the closest point on the diagonal and that the distance Δ-Δ is zero. We refer to a bijection ϕ that minimizes the total squared distance as an *optimal matching* between X and Y. Finding such a matching is a form of the assignment problem, which can be solved with the Hungarian algorithm or the auction algorithm of Bertsekas (Kerber, Morozov & Nigmetov 2017). The L^2-Wasserstein distance $d(X, Y)$ is then the square root of the minimum total squared distance.

Given some probability measure $\rho \subset \mathcal{D}$, we define the Fréchet function

$$F : \mathcal{D} \to \mathbb{R}, \qquad F(Y) = \int_{\mathcal{D}} d(X, Y)^2 \mathrm{d}\rho(X). \tag{4}$$

In the case of a finite sample $S = \{X_i\} \subset \mathcal{D}$, we have $\rho(X) = n^{-1}\delta_S(X)$ and this becomes

$$F(Y) = \frac{1}{n}\sum_{i=1}^{n} d(Y, X_i)^2. \tag{5}$$

A *Fréchet mean* of the sample S is a diagram Y that minimizes $F(Y)$. In general, this is not unique because F can have multiple minimizers. The *Fréchet variance* of S is $F(Y)$. This is a measure of the uncertainty in the sample. If we let ϕ_i: $Y \to X_i$ be an optimal matching of Y with X_i, we can write this as

$$F(Y) = \frac{1}{n}\sum_{i=1}^{n} d(Y, X_i)^2 = \frac{1}{n}\sum_{i=1}^{n}\sum_{y \in Y} \|y - \phi_i(y)\|^2. \tag{6}$$

We can thus attribute a part

$$\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n} \|y - \phi_i(y)\|^2 \tag{7}$$

of the uncertainty to each point $y \in Y$. Unlike the total Fréchet variance, this attribution is again not unique, because there can be multiple optimal matchings. However, the generic case is that the assignment problem does have a unique optimal solution, so we ignore this possibility here.

Given a sample of diagrams, a local minimum of F can be found in finite time (Turner et al. 2014). The mean and variance of a sample can be combined into a *persistence field*, which we discuss further below.

2.8 Persistence fields

In our analysis, we display the statistics of a sample of persistence diagrams $\{X_i\}$ with a *persistence field*, based on a similar but distinct representation proposed by Adams et al. (2017). The goal is to create a visualization of the persistence data that satisfies a number of objectives. The image should

(i) resemble the underlying persistence diagrams,
(ii) reflect the uncertainty in the sample,
(iii) be stable with respect to noise in the data,
(iv) reflect the number of topological features,
(v) show both rare high-persistence features and common low-persistence features.

The first two goals suggest that we use a Fréchet average Y of the sample diagrams (see Section 2.7). This also gives us a measure of the uncertainty σ_y^2 of each feature $y \in Y$. The third and fourth goals suggest some kind of kernel density estimation or smoothing of the average diagram. This creates a difficulty, because those features that are most significant are also extremely rare and are washed out in any kernel density estimate. We therefore assign each feature $y \in Y$ a weight w_y proportional to the square root of its persistence. We then smooth Y with a Tri-cube kernel

$$K(r) = \left(1 - r^3\right)^3, \qquad 0 \leq r \leq 1. \tag{8}$$

The persistence field $f : \mathbb{R}^2 \to \mathbb{R}$ is then

$$f(x) = \sum_{y \in Y} w_y K\left(\|x - y\|/(b\sigma_y)\right), \tag{9}$$

where b is the bandwidth. The Tri-cube kernel has a relatively flat top so that clearly distinguishable features resemble a disc with radius $\sim b\sigma_y$. Our pipeline is illustrated in Fig. 6.

3 STRUCTURAL FILTRATIONS

We study bottom-up filtrations of the ionization bubble network. By considering filtrations along different dimensions (e.g. length-scale or time), we can study different aspects of the ionization topology. Common to all filtrations is the interpretation of the topological features themselves. As explained in Section 1.2, these are the connected ionized regions, the neutral filaments, and the neutral patches enclosed by the bubble network. The topology is characterized in terms of the births and deaths of features at every scale. The precise meaning of this scale, and the interpretation of topological persistence, depends on the dimension along which we build our filtration.

3.1 Spatial structure

First, we consider a snapshot of the ionization field at a fixed redshift. As input, we need the locations $\{x_1, x_2, \dots\}$ of the ionizing sources. We also need to specify the radius r_i of the ionized region surrounding the source at x_i. These data could be the output of a seminumerical model or obtained by applying granulometry (Kakiichi et al. 2017) to the ionization map of a full radiative transfer simulation, or to 21-cm tomographic images (see Section 4). The bubble size distribution could also be constrained by observation through other means (Friedrich et al. 2011; Lin et al. 2016; Giri et al. 2017).

Associate a weight $w_i = r_i$ with the source at x_i. We then use the weighted point set $\mathcal{P} = \{(x_1, w_1), (x_2, w_2), \dots\}$ as the basis for a weighted α-complex. The filtration consists of the (finitely many) distinct α-shapes obtained as we increase the scale from $\alpha = -\infty$ to $\alpha = \infty$. With this filtration, we probe the connectivity structure of the ionization field at a particular redshift. The persistence of a feature has its usual interpretation as topological significance.[8]

This is the only type of filtration that involves both negative and positive values of α. It is worthwhile to pause here and understand why. At $\alpha = 0$, the bubbles have precisely their prescribed radius $\sqrt{r_i^2 + \alpha^2} = r_i$. Negative values of α correspond to deflating the bubbles. A bubble disappears when its deflated radius becomes zero, which happens at $\alpha = -r_i$. Therefore, the bubble size distribution is encoded in the persistent homology of the spatial filtration for negative values of α. Positive values of α correspond to inflating the bubbles. Among other things, this allows us to determine the topological significance of features that exist in the bubble network at $\alpha = 0$ by considering at what scale α_{death} the feature disappears. In Fig. 7, we see the same bubble network at negative, zero, and positive values of α.

Analysing the spatial filtration is particularly useful for investigating the multiscale nature of the largest ionized region that arises as a result of the hierarchical build-up of structure. Small bubbles that have been absorbed into larger bubbles at $\alpha = 0$ must have merged at some $\alpha < 0$. Similarly, clusters that are separated at $\alpha = 0$ will merge when the bubbles are sufficiently inflated, affecting the homology at $\alpha > 0$.

3.2 Bubble dynamics

A second interpretation of the above filtration is obtained by taking cosmic time t as our filtration parameter: $\alpha = t$. Again, we require the source locations $\{x_1, x_2, \dots\}$. Let τ_i be the formation time of the bubble at x_i and define its weight through $w_i = -\tau_i$. We then consider the weighted α-complex with point set $\mathcal{P} = \{(x_1, w_1), (x_2, w_2), \dots\}$. The filtration consists of the distinct α-shapes that we find as we increase time from $t = 0$ to $t = \infty$. The reason for taking negative weights is that it allows us to start at $t = 0$, such that the source at x_i is born when $t = \tau_i$. The bubbles expand when t is increased, simulating the process of reionization. Because weighted α-shapes are based on the distance function (1), this technique requires all bubbles to grow at a non-linear rate $\sim\sqrt{t^2 - \tau_i^2}$ and assumes that the bubbles are spherical.[9] Despite these limitations, this simple toy model already displays many of the qualitative features of reionization. A major conceptual advantage of the α-shape method is therefore that we can take α as a measure of time, allowing us to display the entire topological history of the ionization field in one figure. Moreover, the persistence of a feature can be interpreted as its lifetime.

To circumvent the limitations of the α-shape method, we also propose an alternative method that makes use of field filtrations (Section 2.3). This allows us to consider two further filtrations.

3.3 Ionization gradient

Up until this point, we assumed a binary ionization field and probed the topology along the dimensions of time and space. A third dimension would be the ionization fraction itself. We again start with a time slice of the ionization history, but now build a filtration by taking superlevel sets of the ionization fraction field. This can be done as follows. First, we need a set of vertices $p \in \mathcal{P}$ at which the ionization field is probed. We then construct a Delaunay triangulation $\mathcal{T}$ of $\mathcal{P}$ and a linearly interpolated ionization field with DTFE (Cautun & van de Weygaert 2011). Using these data as input,

[8] When α-shapes are used in pattern recognition, persistence is useful as a criterion for filtering out noise (Edelsbrunner 2010).

[9] In fact, we only require that the bubble network is homotopy equivalent to the spherical network for each value of t, such that the holes coincide. See the discussion in Section 2.1.

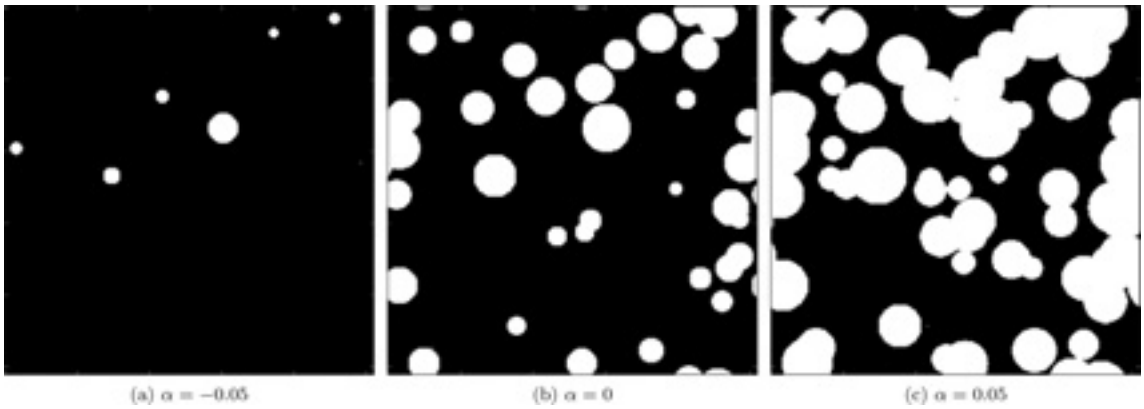


Figure 7. Slices of a non-uniform bubble network with lognormal bubble sizes ($\mu = -3.0$, $\sigma = 0.10$), for different values of α. The median bubble radius is 0.05. This means that at $\alpha = -0.05$, half of all bubbles are *redundant* and have yet to appear. Those that have appeared are deflated. At $\alpha = 0$, all bubbles are present and have exactly their lognormal radius. At $\alpha = 0.05$, the bubbles have been inflated.

we compute the persistent Betti numbers of the field filtration of $\mathcal{T}$. The methodology is essentially the same as in Pranav et al. (2017), except that we replace the matter density field with the ionization fraction field. This method is useful for studying regions that are in the process of being ionized at a particular moment. The persistence of a feature is now interpreted as the differential ionization fraction of the hole.

3.4 Full evolution

Given the output of a more realistic model, we can also build a filtration simply by playing back the ionization history. To do this, we assign every vertex $p \in \mathcal{P}$ a value corresponding to the redshift at which that point was first considered to be part of an ionized region. The filtration is then built by taking superlevel sets of this field. A first goal will be to compare the topology of a full radiative transfer simulation with that of the bubble dynamics model considered in Section 3.2. One caveat that remains is that filtrations are strictly nested sequences, so regions that recombine cannot be handled easily.

4 SOURCE PROPERTIES

In this paper, we study the spatial structure and dynamics of the ionization bubble network using a number of phenomenological models. In each of our models, N sources are placed in a periodic unit box $X \subset \mathbb{R}^3$. The α-shapes are then computed from the set of source locations using the computer package CGAL (Jamin, Pion & Teillaud 2017; The CGAL Project 2017). The topological properties of the network are computed using the algorithm discussed in Section 2.6. We use different methods of generating the bubble locations and weights in order to illustrate different aspects of the reionization process. To this end, we need to specify some of the properties of the ionizing sources.

4.1 Source distribution

The first property that is needed is the spatial distribution of the ionizing sources. In realistic models, the source locations will be correlated with the matter density field. Here, we generate the locations with three spatial point processes. In Section 5.1.2, we use these toy models to demonstrate how the source distribution is reflected in the topology. To isolate the role of the spatial distribution, we assume that all bubbles are born at the same time, in which case the resulting bubble networks are uniformly sized. This means they can be generated with unweighted α-shapes. See Fig. 8 for slices through uniform bubble networks generated according to the different point processes discussed below.

4.1.1 Poisson model

The simplest way to generate the locations is with a Poisson point process with intensity $\Lambda = N$. The actual number of sources is a Poisson random variable, but we tweak the process to ensure that precisely N sources are generated.

4.1.2 Clustered model

The locations in the clustered model are generated with a Neyman–Scott process (Neyman & Scott 1958). The model is described by two parameters K and λ in addition to the number of centres N. Initially, K cluster centres are generated with a Poisson point process. Subsequently, N/K (rounded to the nearest integer) sources are placed with another Poisson process in a sphere of radius $\lambda K^{-1/3}$ around each of the K initial locations. In this way, K clusters of N/K sources are created.

4.1.3 Anticlustered model

The anticlustered model uses a repulsive point process and is described by two parameters: the number of centres N and the minimum centre distance λ. Sources are generated with a Poisson point process and rejected if they fail the minimum distance requirement until N centres have been produced.

4.2 Bubble size

When studying the spatial structure of the bubble network, the size distribution of the ionization bubbles is an important factor. Guided by analytical predictions, many authors have found an approximate lognormal bubble size distribution (Furlanetto et al. 2004a; Furlanetto & Oh 2005; McQuinn et al. 2007; Mesinger &

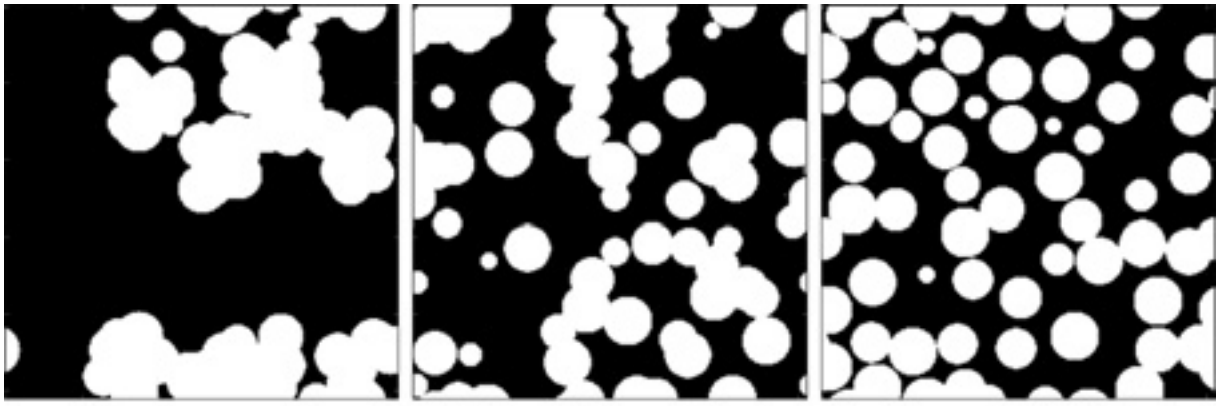


Figure 8. Slices of uniformly sized bubble networks generated with three different point processes. All three pictures correspond to one particular value of $\alpha = 0.06$. Because these are uniform bubble networks, all bubbles have the same radius α.

Furlanetto 2007; Friedrich et al. 2011; Zahn et al. 2011; Lin et al. 2016). The distribution is expected to peak at a characteristic scale that increases and has a variance that decreases as reionization progresses and bubbles merge. This motivates the following phenomenological lognormal model (Coles & Jones 1991).

4.2.1 Lognormal model

In the lognormal model, the source locations x_i are generated with a Poisson process and the bubble sizes r_i are sampled from a lognormal distribution with parameters μ and σ. Fig. 9 shows slices through the resulting bubble networks for different values of (μ, σ). This model is used in Section 5.2.1 to investigate how a changing size distribution is reflected in the topology.

4.2.2 Granulometry

The α-shape method can also be applied to more realistic models of reionization. Since we need to specify bubble centres x_i and radii r_i, we need to find a way to capture the ionized regions in terms of spherical ionization bubbles. One convenient way to do this is with granulometry (Kakiichi et al. 2017), which is based on a mathematically well-defined notion of sieving. Applying this technique to tomographic 21-cm images is a promising pathway for the application of our formalism to observation.

4.3 Bubble age

When we study bubble dynamics, we also need to specify the bubble formation times τ_i. In realistic models, formation times depend on the matter density field and physical properties of reionization, such as the local requirements for source formation. In this case, the spatial distribution of the bubbles and their formation times will be related, affecting the topology of the resulting ionization field.

4.3.1 Constant rate model

In Section 5.1.1, we study the following model in which the number $N_{\rm born}(t)$ of bubbles that have been born at time t increases at a constant rate: $\dot{N}_{\rm born} = \text{const}$ until $t = T$, after which the source production turns off. In this model, the bubble locations x_i are chosen with a Poisson process. Hence, the spatial distribution and formation times are independent. The formation time τ_i of the bubble at x_i is sampled from a uniform distribution $U(0, T)$. As we use weighted α-shapes to model the bubble networks, we set the radius $r_i(t)$ of the bubble with centre x_i at time t equal to

$$r_i(t) = \begin{cases} 0 & \text{if} \quad t < \tau_i, \\ \sqrt{t^2 - \tau_i^2} & \text{otherwise.} \end{cases} \tag{10}$$

This means that the average bubble radius at times $t < T$ will be

$$\langle r(t) \mid \text{alive} \rangle = \int_0^t \frac{\sqrt{t^2 - \tau^2}}{t} \mathrm{d}\tau = \frac{\pi t}{4} \approx 0.785t,$$

whereas the average bubble radius for later times $t \geq T$ is

$$\langle r(t) \rangle = \frac{1}{2T} \left[Tx + t^2 \arctan\left(Tx^{-1}\right) \right] \approx t \qquad \text{for } t \gg T,$$

where $x = \sqrt{t^2 - T^2}$. Hence, the average bubble expands at a rate $\dot{r} = 0.785$ initially, after which it approaches $\dot{r} = 1$ asymptotically. Different trajectories of $\langle r(t) \rangle$ could be effected by sampling τ_i from different distributions. However, our methodology means that we have to use the piecewise function (10).

4.3.2 Physical bubble models

The previous model exhibits the qualitative features of a network of expanding H II bubbles, but uses an unrealistic radius function (10). A more realistic approach would incorporate a physical model of ionization bubbles. Let us briefly discuss two such models.

First, we neglect the effects of cosmological expansion and recombinations. Let the ionizing photon number luminosity of a source be $\dot{N}_\gamma$. Then, the bubble radius at time t is given by

$$r(t) \sim \left[\dot{N}_\gamma (t - \tau) \right]^{1/3}, \tag{11}$$

where τ is again the formation time. This power-law behaviour $r \sim t^{1/3}$ is markedly different from the approximately linear bubble growth $r \sim t$ suggested by equation (10). In both models, the bubble radius grows monotonically and indefinitely due to a lack of recombinations. The most important difference is that the physical model (11) allows for different types of sources with different luminosities $\dot{N}_\gamma$.

To account for recombinations and cosmic expansion, we could instead use the cosmological Strömgen sphere model of Shapiro & Giroux (1987). By specifying a cosmological model, a clumping

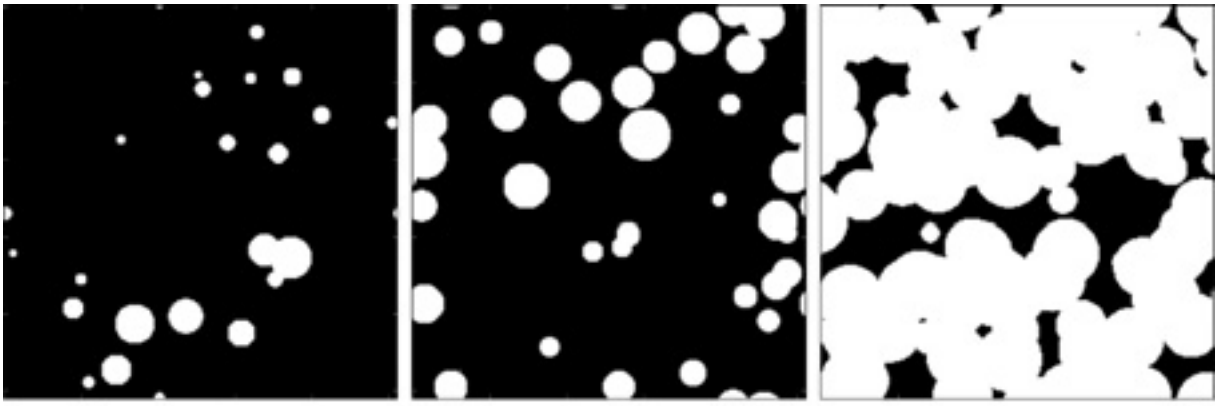


Figure 9. Slices of non-uniform bubble networks generated with lognormal (μ, σ) bubble sizes. All three pictures are taken at $\alpha = 0$, so the bubbles are neither inflated nor deflated. Compare panel 9b with Fig. 7, where the same network is depicted for different α.

factor, and a source function $\dot{N}_\gamma(t)$, this model can be solved for the bubble radius as a function of time. Both the simple physical model (11) and the Shapiro–Giroux model could be implemented using field filtrations or granulometry. However, this comes at the cost of the conceptual simplicity of the α-shape method. We further address these issues in the sequel to this paper.

5 RESULTS

We now use the models of the preceding sections to demonstrate how different aspects of the reionization process affect the homology of bubble network filtrations. In Section 5.1.1, we show how the different stages of reionization can be identified. We then investigate the effect of the spatial distribution of the sources in Section 5.1.2. Finally, we consider the effect of the bubble size distribution in Section 5.2.1.

5.1 Temporal filtrations

We start with a number of temporal filtrations. In Section 5.1.1, we study the constant rate model in which the bubbles are born at a constant rate between $t = 0$ and $t = T$, after which source production is turned off. In the models considered in Section 5.1.2, all bubbles are born at $t = 0$. As the resulting bubble networks are uniformly sized, these latter models could also be interpreted as spatial filtrations.

5.1.1 Stages of reionization

We study the different stages of reionization with a temporal filtration of the constant rate model with $N = 500$ bubbles and $T = 0.10$. The results, averaged over 10 realizations, are shown in Fig. 10. In the top panel, we see the Betti curves describing the number of ionization bubbles (Betti-0, solid black), neutral filaments (Betti-1, solid red), and neutral patches (Betti-2, dashed blue). We have overlaid the global ionization fraction $Q(t)$, which is the fraction of total volume occupied by ionization bubbles (dot–dashed). The number of ionized regions β_0 starts at 0 and initially just tracks the number

$$N_{\rm born}(t) = \frac{Nt}{T}$$

of bubbles that have been born (dotted line). After a while, the slope of the β_0-curve tapers off as bubbles start to overlap and merge. We can therefore use β_0 as a measure of the degree of overlap. At $t = 0.038$, the degree of overlap $1 - \beta_0/N_{\rm born}$ has reached 10 per cent. This could be chosen as the end of the pre-overlap stage. Soon after this point, β_0 reaches a maximum despite the fact that new bubbles are still being created. Notice that the ionization fraction $Q(t)$ only starts to incline appreciably in the subsequent overlap stage. This is confirmed by the inset graph, which shows the Betti numbers as a function of Q.

During the overlap stage, the β_0-curve never reaches far above 200 because any newborn bubbles are immediately fed into larger existing structures. Furthermore, because new bubbles are created at a constant rate up to $t = 0.10$, the β_0-curve is skewed very much to the right and has a long and fat tail. At $t = 0.087$, a percolation transition occurs and one large connected ionized region now stretches from one side of the simulation box to the other. At this point, the degree of overlap has increased to 94 per cent, signalling the end of the overlap stage. Strikingly, it is at this point also that the topology starts to become highly filamentary, which agrees with the findings of Bag et al. (2018). At the transition, the volume ionization fraction has a value of $Q = 0.26$. This is significantly larger than the values found by Furlanetto & Oh (2016) and Bag et al. (2018), which can be explained by the fact that bubble locations are uncorrelated in this model.

Soon after percolation, the ionization rate reaches a maximum at $t = 0.097$. During the post-overlap stage, the remaining neutral islands are attacked from the outside. Interestingly, most of the higher dimensional structure only appears past this point. First, the number β_1 of tunnels increases as bubbles begin to overlap that were already connected, forming 1-cycles. These tunnels surround neutral filaments that pierce through the ionization bubble network. When the filaments are ionized and tunnels begin to be filled up, β_1 decreases. Meanwhile, β_2 increases as bubbles start to enclose an increasing number of neutral patches. After $t = 0.126$, the patches outnumber the tunnels. Finally, the patches get ionized as well and the Betti numbers reach their final values: $\beta_0 = 1$, $\beta_1 = 3$, $\beta_2 = 3$. These values are an artefact of the non-trivial topology of the periodic simulation box X, but are negligible compared to the dozens of features found at earlier times.

Looking at the persistence diagrams in the bottom row of Fig. 10, we see that the majority of features are short lived (close to the diagonal). Nevertheless, a large number of tunnels that are born

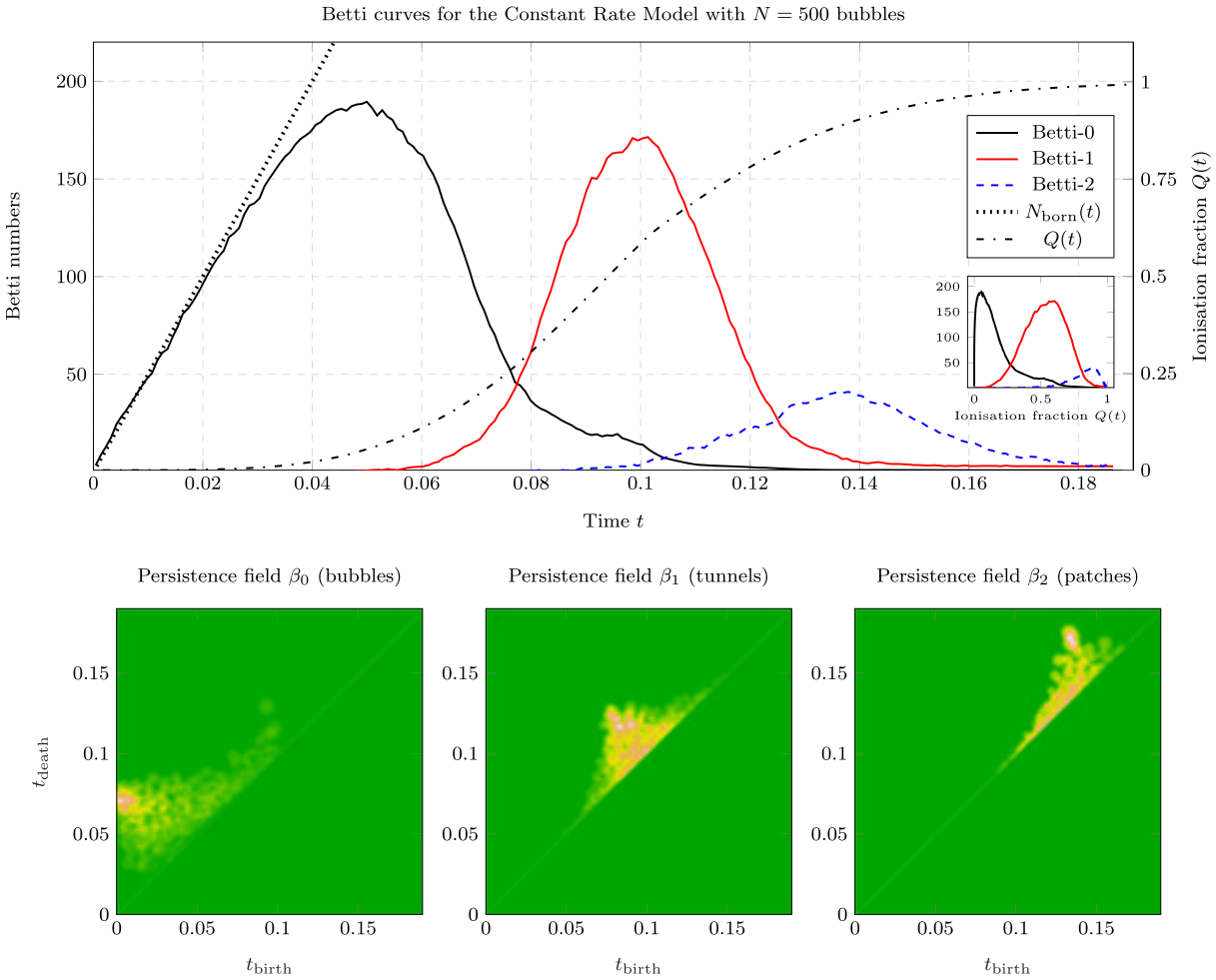


Figure 10. Persistent homology of the constant rate model with $N = 500$ bubbles. In the top panel, we see the number of ionization bubbles (β_0), neutral filaments (β_1), and neutral patches (β_2) alive at any time t. Also shown is the number $N_{\rm born}(t)$ of bubbles that have been born and the global ionization fraction $Q(t)$. The inset shows the evolution of the Betti numbers as a function of the ionization fraction. In the bottom panels, we see the persistence fields showing the births and deaths of all topological features.

Table 1. Different epochs in the $N = 500$ constant rate model.

Epoch	Ends when	Time
Pre-overlap	10 per cent of bubbles overlap	$t = 0.038$
Overlap	Percolation occurs	$t = 0.087$
Filament	Patches outnumber tunnels	$t = 0.126$
Patch	Reionization is complete	$t = 0.186$

around $t = 0.07$ survive until $t = 0.10$, although none live past $t = 0.13$. Furthermore, most of the patches that are born before $t = 0.13$ die very young, but a large number of patches that are born at $t = 0.14$ survive until $t = 0.17$. We thus identify two additional topologically significant epochs past the pre-overlap stage and the overlap stage, which may rightly be called the 'filament stage' and 'patch stage'. These topological characteristics are not apparent from the geometry or ionization history $Q(t)$. We summarize our criteria for the four stages in Table 1. We also have a persistence field for β_0, which shows that the longest-living ionized regions emerge at $t = 0$, though even at $t = 0.10$ some bubbles are born that survive for a relatively long time before being absorbed into larger structures.

It is worth noting that our choice to end the overlap stage at the point of the percolation transition differs from Furlanetto & Oh (2016), who identify the percolation transition as the division between the pre-overlap and overlap stages. The reason for our convention is that the topology has a distinct character during each of the stages. During the pre-overlap stage, the topology is characterized by the birth of localized bubbles with little to no overlap (less than 10 per cent). The overlap stage is characterized by the growth of ever larger connected structures, ending with the percolation transition. The network then enters a filament stage during which the remaining highly filamentary clusters merge. In the final stage, the topology is dominated by enclosed neutral patches.

5.1.2 *Source distribution*

To illustrate how the source distribution affects the topology, we study three different models with $N = 500$ uniformly sized bubbles placed at random locations: a clustered model, an anticlustered model, and a Poisson model. A visual inspection of the resulting bubble networks shown in Fig. 8 is quite revealing. We see that the Poisson model is an intermediate case between two extremes. The bubble network produced by the clustered model resembles a two-

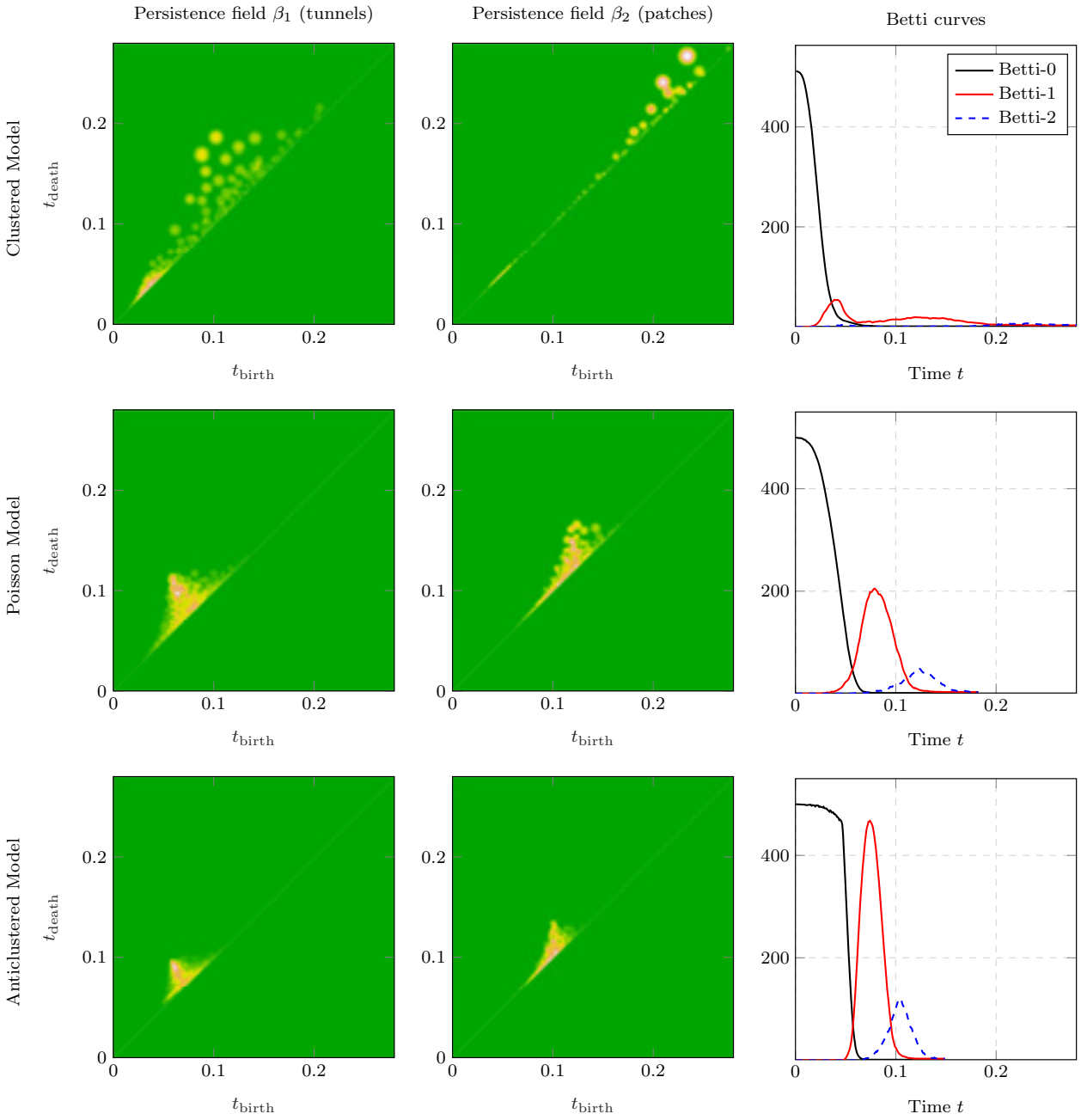


Figure 11. The impact of clustering on persistent homology. The features in the clustered model (top) are rare, but far more persistent and spread out compared to the anticlustered model (bottom). The Poisson model (middle) is an intermediate case. Notice also that there are two generations of features in the clustered model. The early low-persistence features correspond to mergers within clusters and the late high-persistence features correspond to mergers of clusters.

phase medium consisting of large clusters of bubbles and neutral oceans utterly devoid of bubbles. As a result, the clustered regions are rapidly ionized, but the neutral regions resist ionization for a long time. At the other extreme, the anticlustered model produces bubbles appearing in an almost crystal-like pattern. For a long time, these bubbles can freely expand in every direction and when the bubbles finally overlap, the box is almost completely ionized.

To produce these bubble networks, we choose rather extreme model parameters. For the clustered model, we generate $K = 32$ superclusters with a characteristic size $\lambda K^{-1/3} = 0.08$ with $\lambda = 0.25$. For the anticlustered model, we place the sources at least a distance $\lambda = 0.094$ from any other source. The results averaged over 10 realizations are shown in Fig. 11. The differences are conspicuous. First, consider the Betti curves in the third column. In contrast to the constant rate model, the β_0-curves all start at 500 because all bubbles are born simultaneously. Looking at the top right-hand panel, it appears as if the patch stage is absent in the clustered model. At the height of the patch epoch, there are only $\beta_2 = 7.0$ neutral patches on average, compared to $\beta_2 = 48.0$ patches in the Poisson model. This is because during the patch epoch, the clustered regions

are completely filled up and lack any tunnels or patches. The only remaining patches are the huge empty bubble-less regions, which are few in number but large in size. This is obvious when we consider the persistence diagrams for patches in the second column.

In the bottom right-hand panel, we see that the anticlustered model has a very long pre-overlap stage during which the number β_0 of ionized regions plateaus. Because the minimum bubble separation λ was set rather high, most centres have a closest neighbour at a distance of roughly λ. Therefore, the bubble network goes through a swift phase transition at $t = \frac{1}{2}\lambda = 0.047$, when the bubbles start to overlap. We also see that the anticlustered model has a more significant patch epoch and a brief but extreme filament epoch. The number of tunnels almost reaches 500, nicely adhering to the crystalline expectation.

Looking at the persistence diagrams in the first two columns, we see that although the clustered model has fewer higher dimensional structures, they are far more persistent and appear over a much wider time interval. For the anticlustered model, we see that there are many more tunnels and patches, but they exist only during a very short period of time. Again, we find that the apparent intensities of the tunnel and patch epochs in the Betti diagrams are deceiving: the clustered model does have a patch epoch, but there are fewer yet more significant patches. The opposite is true for the anti-clustered model. The Poisson model is once more an intermediate case.

The persistence diagrams of the clustered model in the top row show additional structure. Observe that there are two distinct generations of features. This is a reflection of the fractal-like multiscale topology produced by the Neyman–Scott process. The small-scale low-persistence features correspond to structure that emerges early on within clusters. The large-scale high-persistence generation consists of global features that arise when clusters merge with clusters.

For each of the models, we again find that a percolation transition occurs around the point where the topology becomes dominated by filaments. The corresponding ionization fraction is $Q = 0.225$ for the clustered model, $Q = 0.250$ for the Poisson model, and $Q = 0.348$ for the anticlustered model. This is in line with the expectation that correlation between the bubble locations induces a percolation transition at lower ionization fractions (Furlanetto & Oh 2016).

5.2 Spatial filtrations

We now consider filtrations of bubble networks with a given size distribution. These are spatial filtrations, which means that the interpretation is somewhat different from the temporal filtration described above. Refer to Section 3 for a discussion of these differences. The filtration parameter is the scale α.

5.2.1 Bubble size distribution

To illustrate these ideas, we consider three models with a lognormal size distribution with mean μ and standard deviation σ. The average bubble radius is

$$\langle R \rangle = \mathrm{e}^{\mu+\sigma^2/2}.$$

The results for $N = 500$ bubbles, again averaged over 10 realizations, are shown in Fig. 12. The bubbles are smallest in the top row ($\mu = -4.0$, $\sigma = 0.50$), largest in the bottom row ($\mu = -2.5$, $\sigma = 0.05$), with the middle row being an inbetween case ($\mu = -3.0$, $\sigma = 0.10$).

The first two columns show the persistence fields for ionized bubbles (β_0) and neutral filaments (β_1). We have divided the persistence fields into quadrants. Features in quadrant II ($\alpha_{\mathrm{birth}} \leq 0$, $\alpha_{\mathrm{death}} > 0$) are present in the bubble network. Features in quadrant I exist only over positive scales. The Betti curves in the third column indicate the total numbers of features alive at every scale. The homology of the bubble network can be read off by looking at the intersections of the Betti curves with the line $\alpha = 0$.

We have overlaid the lognormal probability density functions on the β_0 persistence fields. By construction, the scales α_{birth} at which bubbles are born follow the lognormal distribution. In the top row, we see that the largest bubbles die at large scales. But many of the smallest bubbles die at small scales. This is because of the elder rule, which states that whenever two features merge, the oldest feature survives. We have also overlaid the cumulative distribution functions $P(R \geq -\alpha)$ on the Betti curves. The β_0-curve follows the distribution function when α is small, but deviates once bubbles start to merge.

We can identify what stage of the reionization process the bubble network is currently undergoing by considering the distribution of features over the quadrants:

(i) When the bubbles are smallest (top), almost all features in the β_0-field are in quadrant II and all features in the β_1-field are in quadrant I. We conclude that the topology is dominated by separated islands of ionized material. The bubble network is in the pre-overlap stage.

(ii) In the second row, about a third of the zero-dimensional features are in quadrant II and two thirds are in quadrant III. Most of the tunnels are in quadrant I, although a few are in quadrants II and III. The bubble network is therefore in the overlap stage, and entering the filament stage.

(iii) Finally, when the bubbles are largest (bottom), all zero-dimensional features are in quadrant III. This means that they have all merged into a single connected component. A preponderance of the tunnels are in quadrant II and therefore alive in the bubble network. We are in the filament stage.

In each case, a quick glance at the bubble network slices shown in Fig. 9 confirms the picture sketched by the division of features over the quadrants. Overall, the Betti curves and persistence fields most resemble the Poisson model. This is not surprising, because the source locations were generated with a Poisson point process. Finally, note that the β_1-fields look like translated and distorted copies of each other. The reason that the fields are not just translated copies is the non-linear bubble scaling $\sim\sqrt{R^2 + \alpha^2}$.

6 DISCUSSION

The formalism presented here provides a substantial deepening of our understanding of the topology of H II regions during the Epoch of Reionization. Homology allows us to characterize the topology of the ionization bubble network in terms of its components (ionized regions), tunnels (enclosed neutral filaments), and cavities (neutral patches), collectively called topological features. Persistence is a measure of the significance of a feature. We have shown that the persistence of a feature can be variously interpreted as its lifetime, significance, or ionization state. Persistent homology provides us with quantitative measures that are more general than other commonly used measures such as the Euler characteristic, Minkowski functionals, and the bubble size distribution.

We use the tool of α-shapes, borrowed from computational topology, to model the ionization bubble network. Together with

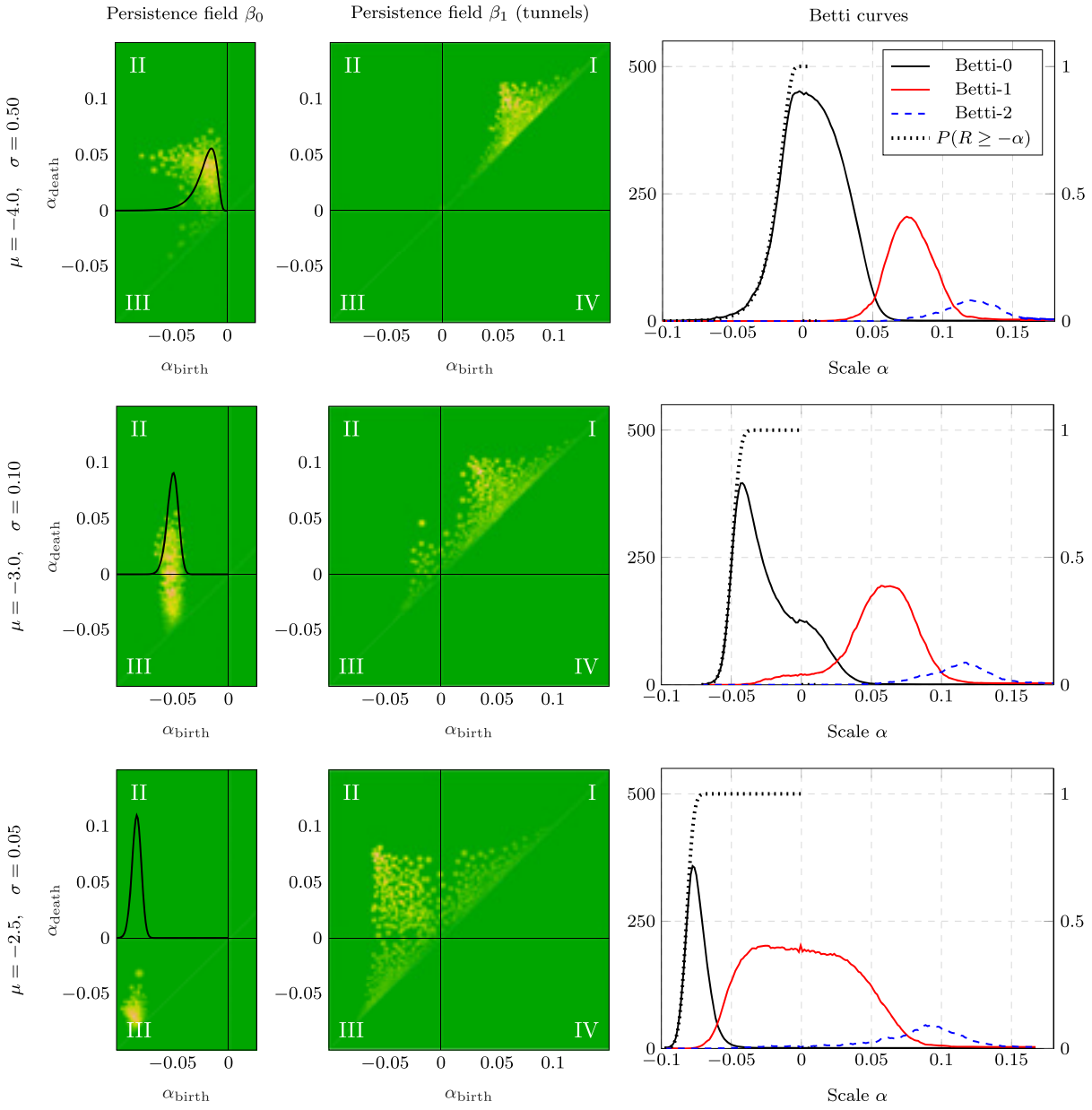


Figure 12. Persistent homology of a bubble network with lognormally distributed bubble sizes, for different values of μ and σ. Bubbles are smallest for $\mu = -4.0$ (top) and largest for $\mu = -2.5$ (bottom). The lognormal probability density is overlaid on the β_0-persistence fields (left). The cumulative distribution function is overlaid on the Betti curves (right). Features in quadrant II of the persistence fields are alive at $\alpha = 0$.

persistent homology, α-shapes are ideally suited to study every stage of the reionization process. Starting at the pre-overlap and overlap stages, we can follow the number of distinct ionized regions as bubbles arise and subsequently merge. During the later stages, homology allows us to understand the topology of the bubble network as a large fractal-like structure, pierced by neutral filaments and enclosing patches of neutral hydrogen. The topology of the bubble network depends on the underlying physics through an interaction of the spatial distribution of the ionizing sources with the size distribution of the surrounding H II regions.

This work is a stepping stone for a number of further studies on the persistent topology of reionization. In an upcoming paper, we apply our methods to a physical model of reionization. Ultimately, the goal is to study the topology of reionization using 21-cm observations. As illustrated with the phenomenological models in this paper, what is needed is a specification of the spatial and size distribution of H II regions. A first step will be to study the viability of sufficiently constraining these properties with upcoming observations. Further statistical analysis will be necessary to determine the requirements on future experiments that would definitively allow a homological study of reionization. However, the statistics of persistence diagrams has only recently been put on firm footing, so more work is needed on this front. Finally, we have proposed a more general filtration method that relaxes the α-shape assumptions. This method would enable us to study the evolving topology of a fully dynamic ionization fraction field, given a set of 3D measurements of the ionization field at multiple redshifts.

ACKNOWLEDGEMENTS

RvdW is grateful for numerous useful, instructive, and insightful discussions with Gert Vegter, Bernard Jones, Job Feldbrugge, Garrelt Mellema, and Keimpe Nevenzeel. WE similarly thanks Martijn Oei and Kees Elbers for helpful discussions. We also thank the anonymous referee, whose comments helped to improve the manuscript.

REFERENCES

Adams H. et al., 2017, J. Mach. Learn. Res., 18, 218
Bag S., Mondal R., Sarkar P., Bharadwaj S., Sahni V., 2018, MNRAS, 477, 1984
Beardsley A. P. et al., 2016, ApJ, 833, 102
Cautun M. C., van de Weygaert R., 2011, Astrophysics Source Code Library, record ascl:1105.003
Chardin J., Aubert D., Ocvirk P., 2012, A&A, 548, A9
Choudhury T. R., Haehnelt M. G., Regan J., 2009, MNRAS, 394, 960
Cole A., Shiu G., 2018, J. Cosmol. Astropart. Phys., 2018, 025
Coles P., Jones B., 1991, MNRAS, 248, 1
Dixon K. L., Iliev I. T., Mellema G., Ahn K., Shapiro P. R., 2016, MNRAS, 456, 3011
Doussot A., Trac H., Cen R., 2019, ApJ, 870, 18
Edelsbrunner H., 1992, Weighted Alpha Shapes, Vol. 92. University of Illinois at Urbana–Champaign, Illinois
Edelsbrunner H., 2010 , in van de Weygaert R., Vegter G., Ritzerveld J., Icke V., eds, Tesselations in the Sciences; Virtues, Techniques and Applications of Geometric Tilings. Spinger-Verlag, Heidelberg, p. 1
Edelsbrunner H., Mücke E. P., 1994, ACM Trans. Graph., 13, 43
Edelsbrunner H., Kirkpatrick D., Seidel R., 1983, IEEE Trans. Inf. Theory, 29, 551
Edelsbrunner H., Facello M., Fu P., Liang J., 1995, in Hunter L., ed., Proceedings of the Twenty-Eighth Annual Hawaii International Conference on System Sciences, IEEE, New York, p. 256
Edelsbrunner H., Letscher D., Zomorodian A., 2002, Discrete Comput Geom, 28, 511
Feldbrugge J., van Engelen M., 2012, BSc thesis, Univ. Groningen
Feldbrugge J., van Engelen M., van de Weygaert R., Vegter G., 2018, MNRAS, submitted
Finlator K., Özel F., Davé R., Oppenheimer B. D., 2009, MNRAS, 400, 1049
Friedrich M. M., Mellema G., Alvarez M. A., Shapiro P. R., Iliev I. T., 2011, MNRAS, 413, 1353
Furlanetto S. R., Oh S. P., 2005, MNRAS, 363, 1031
Furlanetto S. R., Oh S. P., 2016, MNRAS, 457, 1813
Furlanetto S., Hernquist L., Zaldarriaga M., 2004a, MNRAS, 354, 695
Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004b, ApJ, 613, 1
Giri S. K., Mellema G., Dixon K. L., Iliev I. T., 2017, MNRAS, 473, 2949
Gleser L., Nusser A., Ciardi B., Desjacques V., 2006, MNRAS, 370, 1329
Gnedin N. Y., 2000, ApJ, 535, 530
Gnedin N. Y., 2014, ApJ, 793, 29
Hatcher A., 2002, Algebraic Topology, Cambridge University Press, Cambridge
Hong S. E., Ahn K., Park C., Kim J., Iliev I. T., Mellema G., 2014, J. Korean Astron. Soc., 47, 49
Hutter A., 2018, MNRAS, 477, 1549
Icke V., van de Weygaert R., 1987, A&A, 184, 16
Iliev I. T., Mellema G., Pen U.-L., Merz H., Shapiro P. R., Alvarez M. A., 2006, MNRAS, 369, 1625
Iliev I. T., Mellema G., Ahn K., Shapiro P. R., Mao Y., Pen U.-L., 2014, MNRAS, 439, 725
Jamin C., Pion S., Teillaud M., 2017, 3D Triangulations, 4.10 edn. CGAL Editorial Board
Kakiichi K. et al., 2017, MNRAS, 471, 1936
Kapahtia A., Chingangbam P., Appleby S., Park C., 2018, J. Cosmol. Astropart. Phys., 2018, 11
Katz H., Kimm T., Haehnelt M. G., Sijacki D., Rosdahl J., Blaizot J., 2018, MNRAS, 483, 1029
Kerber M., Morozov D., Nigmetov A., 2017, J. Exp. Algorithmics, 22, 1
Kerrigan J. R. et al., 2018, ApJ, 864, 131
Lee K.-G., Cen R., Gott J. R., III, Trac H., 2008, ApJ, 675, 8
Lin Y., Oh S. P., Furlanetto S. R., Sutter P. M., 2016, MNRAS, 461, 3361
Majumdar S., Mellema G., Datta K. K., Jensen H., Choudhury T. R., Bharadwaj S., Friedrich M. M., 2014, MNRAS, 443, 2843
Makarenko I., Shukurov A., Henderson R., Rodrigues L. F. S., Bushby P., Fletcher A., 2018, MNRAS, 475, 1843
Malloy M., Lidz A., 2013, ApJ, 767, 68
McQuinn M., Lidz A., Zahn O., Dutta S., Hernquist L., Zaldarriaga M., 2007, MNRAS, 377, 1043
Mellema G., Koopmans L., Shukla H., Datta K. K., Mesinger A., Majumdar S., 2015, in Advancing Astrophysics with the Square Kilometre Array. p. 010
Mesinger A., Furlanetto S., 2007, ApJ, 669, 663
Mesinger A., Furlanetto S., Cen R., 2011, MNRAS, 411, 955
Mileyko Y., Mukherjee S., Harer J., 2011, Inverse Probl., 27, 124007
Munkres J. R., 1984, Elements of Algebraic Topology, Vol. 2, . Addison-Wesley, Menlo Park
Nevenzeel K., 2013, MSc thesis, Univ. Groningen
Neyman J., Scott E. L., 1958, J. R. Stat. Soc. B, 20, 1
Ocvirk P. et al., 2016, MNRAS, 463, 1462
Okabe A., 1992, Spatial Tessellations. Wiley Online Library
Park C. et al., 2013, J. Korean Astron. Soc., 46, 125
Patil A. H. et al., 2017, ApJ, 838, 65
Pawlik A. H., Rahmati A., Schaye J., Jeon M., Dalla Vecchia C., 2017, MNRAS, 466, 960
Platen E., van de Weygaert R., Jones B. J. T., 2007, MNRAS, 380, 551
Pranav P., Edelsbrunner H., van de Weygaert R., Vegter G., Kerber M., Jones B. J. T., Wintraecken M., 2017, MNRAS, 465, 4281
Pranav P., van de Weygaert R., Vegter G., Jones B., Adler R., Feldbrugge J., Park C., Buchert T., Kerber M., 2018, MNRAS, 485, 4167,
Pritchard J. R., Loeb A., 2012, Rep. Prog. Phys., 75, 086901
Schaap W. E., van de Weygaert R., 2000, A&A, 363, L29
Shandarin S. F., Sheth J. V., Sahni V., 2004, MNRAS, 353, 162
Shapiro P. R., Giroux M. L., 1987, ApJ, 321, L107
Sheth J. V., Sahni V., Shandarin S. F., Sathyaprakash B. S., 2003, MNRAS, 343, 22
Shimabukuro H., Yoshiura S., Takahashi K., Yokoyama S., Ichiki K., 2017, MNRAS, 468, 1542
Sousbie T., 2011, MNRAS, 414, 350
The CGAL Project, 2017, CGAL User and Reference Manual , 4.10 edn, CGAL Editorial Board
Turner K., Mileyko Y., Mukherjee S., Harer J., 2014, Discrete Comput. Geom., 52, 44
van de Weygaert R., 1994, A&A, 283, 361
van de Weygaert R., Schaap W., 2008, in Martínez V. J., Saar E., Martínez-González E., Pons-Bordería M.-J., eds, Lecture Notes in Physics, Vol. 665. Data Analysis in Cosmology. Springer-Verlag, Berlin, p. 291
van de Weygaert R., et al., 2011, in Gavrilova M. L., Tan C. K., Mostafavi M. A., eds, Transactions on Computational Science XIV. Springer-Verlag, Heidelberg, p. 60
Wasserman L., 2018, Annu. Rev. Stat. Appl., 5, 501
Xu X., Cisewski-Kehe J., Green S. B., Nagai D., 2019, Astron. Comput., 27, 34
Yoshiura S., Shimabukuro H., Takahashi K., Matsubara T., 2017, MNRAS, 465, 394
Zahn O., Mesinger A., McQuinn M., Trac H., Cen R., Hernquist L. E., 2011, MNRAS, 414, 727
Zomorodian A., 2012, Advances in Applied and Computational Topology, American Mathematical Society, Providence, Rhode Island
Zomorodian A., Carlsson G., 2005, Discrete Comput. Geom., 33, 249

This paper has been typeset from a TEX/LATEX file prepared by the author.